Robust covariance matrix estimates with attractive asymptotic properties

M. Mahot and F. Pascal
SONDRA, Supelec
3 rue Joliot-Curie
91190 Gif-sur-Yvette, France
Email: melanie.mahot@supelec.fr
Email: frederic.pascal@supelec.fr

P. Forster
SATIE, ENS Cachan,
CNRS, UniverSud
61, Av. du Pdt Wilson,
F-94230 Cachan, France
Email: philippe.forster@satie.ens-cachan.fr

J. P. Ovarlez
ONERA, DEMR/TSI
Chemin de la Hunière,
91761 Palaiseau
Cedex, France
Email: jean-philippe.ovarlez@onera.fr

***Abstract*—The Sample Covariance Matrix (SCM) is widely used in signal processing applications which require the estimation of the data covariance matrix. Indeed it exhibits good statistical properties and tractability. However its performance can become very bad in context of non-Gaussian signals, contaminated or missing data. In that case M-estimators provide a good alternative. They have been introduced within the framework of elliptical distributions which encompass a large number of well-known distributions as for instance the Gaussian, the K-distribution or the multivariate Student (or t) distribution. In this paper, we show that with an appropriate normalization, the SCM and M-estimators have the same asymptotic behavior. More precisely, they share the same asymptotic covariance up to a scale factor. Tyler (1983) obtains similar results but we propose here a simpler proof for the case of M-estimators. The important consequence is that the SCM can easily be replaced by M-estimators with minor changes in performance analysis of signal processing algorithms. This result is highlighted by simulations in Direction-Of-Arrival (DOA) estimation using a MUltiple SIgnal Classification (MUSIC) approach. In this paper, we address the case of real data. These results have also been extended to the complex case but, due to the lack of space and for clarity of the presentation, this generalization will be omitted and will be addressed later.**

I. INTRODUCTION

Many signal processing applications require the estimation of the data covariance matrix. In the signal processing community, the data are traditionally considered to be Gaussian and the standard covariance matrix estimate is the well-known SCM. However, the SCM suffers from major drawbacks. Firstly, when the data turn out to be non-Gaussian, as for instance in adaptive Radar and Sonar [1], the SCM is a bad estimate: indeed, it is very sensitive to large data and performs poorly in the case of impulsive noise. Secondly, it is not robust to outliers. To overcome these problems, there has been an intense research activity in robust estimation theory in the statistical community these last decades [2], [3], [4]. Among several solutions, the so-called M-estimators originally introduced by Huber [5] and investigated in the seminal work of Maronna [6], have imposed themselves as an appealing alternative to the classical SCM. They have been introduced within the framework of elliptical distributions. Elliptical distributions (for details see [7] chapter 13) encompass a large number of well-known distributions as for instance the Gaussian, the K-distribution or the multivariate Student (or t) distribution. They may also be used to model heavy tailed distributions, as may be met for instance in adaptive Radar with impulsive clutter. M-estimators of the covariance matrix are however seldom used in the signal processing community. Notable exceptions are the recent papers by Ollila [8], [9], [10] who advocates their use in several applications such as array processing. We may also mention for instance, papers in the Spherically Invariant Random Vectors (SIRV) framework for adaptive radar detection [11]. One possible reason for this lack of interest is that their statistical properties are not well-known in the signal processing community, as opposed to the Wishart distribution of the SCM in the Gaussian context. To promote the use of M-estimators, we show in this paper that their asymptotic distribution is essentially the same as the Wishart one. Tyler in [12] obtains similar results but we propose here a simpler proof for the case of M-estimators. More precisely, this result is valid after an appropriate normalization of the covariance matrix estimate, which can be introduced in most signal processing applications without any consequence on final results. For instance, in Direction-of-Arrival (DOA) estimation, a scale factor on the covariance matrix estimate has no influence on the estimated DOAs. In this paper, we address the case of real data. These results have also been extended to the complex case but, due to the lack of space and for clarity of the presentation, this generalization will be omitted and will be addressed later.

This paper is organized as follows. Section II introduces the required background on M-estimators while Section III provides our contribution on the asymptotic distribution of these estimators. Then, in Section IV, simulations validate the theoretical analysis and Section V concludes this work.

Vectors (resp. matrices) are denoted by bold-faced lowercase letters (resp. uppercase letters). $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Lambda})$ denotes the multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Lambda}$. $\sim$ means "distributed as", $\overset{d}{=}$ stands for "shares the same distribution as", $\overset{d}{\to}$ denotes convergence in distribution and $\otimes$ denotes the kronecker product.

II. BACKGROUND

A. *Elliptical distribution*

Let $\mathbf{x}$ be a m-dimensional real random vector. $\mathbf{x}$ has an elliptical distribution if its probability density function (PDF)

978-1-4577-2105-2/11/$26.00 ©2011IEEE

can be written as

$$f_{\mathbf{x}}(\mathbf{x}) = |\mathbf{\Lambda}|^{-1/2} g((\mathbf{x}-\boldsymbol{\mu})^T \mathbf{\Lambda}^{-1}(\mathbf{x}-\boldsymbol{\mu})), \quad (1)$$

where $g : [0, \infty) \to [0, \infty)$ is any function such that (1) defines a PDF, $\boldsymbol{\mu}$ is the statistical mean and $\mathbf{\Lambda}$ is a scatter matrix. The scatter matrix $\mathbf{\Lambda}$ reflects the structure of the covariance matrix of $\mathbf{x}$, i.e. the covariance matrix is equal to $\mathbf{\Lambda}$ up to a scale factor. This elliptical distribution will be denoted by $E(\boldsymbol{\mu}, \mathbf{\Lambda})$.

In this paper, we will assume without loss of generality that *the scatter matrix is equal to the covariance matrix.* Indeed, function g in (1) can always be defined such that this equality holds.

B. M-estimator of the scatter matrix

Let $(\mathbf{x}_1, ..., \mathbf{x}_N)$ be a N-sample of m-dimensional real independent vectors with $\mathbf{x}_i \sim E(\mathbf{0}, \mathbf{\Lambda})$, $i = 1, ..., N$. The M-estimator of $\mathbf{\Lambda}$ is defined as the solution of the following equation

$$\mathbf{V}_N = \frac{1}{N} \sum_{n=1}^{N} u\left(\mathbf{x}_n^T \mathbf{V}_N^{-1} \mathbf{x}_n\right) \mathbf{x}_n \mathbf{x}_n^T, \quad (2)$$

where u is a function satisfying a set of general assumptions stated in [6] and recalled here below in the case where $\boldsymbol{\mu} = \mathbf{0}$:

- u is non-negative, non increasing, and continuous on $[0, \infty)$.
- Let $\psi(s) = su(s)$ and $K = \sup_{s \geq 0} \psi(s)$. $m < K < \infty$, ψ is increasing, and strictly increasing on the interval where $\psi < K$.
- Let $P_N(.)$ denote the empirical distribution of $(\mathbf{x}_1, ..., \mathbf{x}_N)$. There exists $a > 0$ such that for every hyperplane H, $\dim(H) \leq m-1$, $P_N(H) \leq 1 - \frac{m}{K} - a$. This assumption can be strongly relaxed as shown in [13], [14].

Let us now consider the following equation, which is roughly speaking the limit of (2)

$$\mathbf{V} = E\left[u(\mathbf{x}^T \mathbf{V}^{-1} \mathbf{x})\, \mathbf{x}\mathbf{x}^T\right], \quad (3)$$

where $\mathbf{x} \sim E(\mathbf{0}, \mathbf{\Lambda})$. Then, under the above conditions, Maronna has shown in [6] that:

- Equation (3)(resp. (2)) admits a unique solution $\mathbf{V}$ (resp. $\mathbf{V}_N$) and
$$\mathbf{V} = \sigma \mathbf{\Lambda}, \quad (4)$$
where σ is given in [7].
- A simple iterative procedure provides $\mathbf{V}_N$.
- $\mathbf{V}_N$ is a consistent estimate of $\mathbf{V}$.
- The asymptotic distribution of $\mathbf{V}_N$ is given by [15]
$$\sqrt{N}\left(\text{vec}(\mathbf{V}_N) - \text{vec}(\mathbf{V})\right) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \mathbf{\Pi}\right), \quad (5)$$
where $\mathbf{\Pi} = \sigma_1(\mathbf{I}+\mathbf{K})(\mathbf{V}\otimes\mathbf{V}) + \sigma_2(\text{vec}\mathbf{V})(\text{vec}\mathbf{V})^T$, $\mathbf{K}$ is the commutation matrix which transforms $\text{vec}(\mathbf{A})$ into $\text{vec}(\mathbf{A}^T)$, σ_1 and σ_2 are scalars given in [15], [7].

C. Wishart distribution

The Wishart distribution $W(N, \mathbf{\Lambda})$ is the well known distribution of $\sum_{n=1}^{N} \mathbf{y}_n \mathbf{y}_n^T$, where $\mathbf{y}_n$ are i.i.d and $\mathcal{N}(\mathbf{0}, \mathbf{\Lambda})$ distributed. Let $\mathbf{W}_N = N^{-1} \sum_{n=1}^{N} \mathbf{y}_n \mathbf{y}_n^T$ be the related SCM which will be also referred to, with a slight abuse, as a Wishart matrix. The asymptotic distribution of the Wishart matrix $\mathbf{W}_N$ is

$$\sqrt{N}(\text{vec}(\mathbf{W}_N) - \text{vec}(\mathbf{\Lambda})) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \mathbf{\Sigma}\right), \quad (6)$$

where $\Sigma = (\mathbf{I}+\mathbf{K})(\mathbf{\Lambda}\otimes\mathbf{\Lambda})$.

III. Asymptotic distribution of normalized M-estimators of the scatter matrix

A. Main results

The aim of this paper is to show that *after an appropriate normalization* defined here below, M-estimators and Wishart matrices share the same asymptotic distribution. The required normalization is as follows. Let $\mathbf{A}$ be a $m \times m$ matrix. The $\mathbf{\Lambda}$-normalized matrix $\widetilde{\mathbf{A}}$ is defined as

$$\widetilde{\mathbf{A}} = \frac{m}{\text{tr}(\mathbf{\Lambda}^{-1}\mathbf{A})}\, \mathbf{A}. \quad (7)$$

Note that, for any matrix $\mathbf{A} = \alpha\mathbf{\Lambda}$, we have $\widetilde{\mathbf{A}} = \mathbf{\Lambda}$ and therefore:

$$\widetilde{\mathbf{\Lambda}} = \mathbf{\Lambda} \text{ and } \widetilde{\mathbf{V}} = \mathbf{\Lambda}. \quad (8)$$

Finally, using the identity $\text{tr}(\mathbf{A}^T\mathbf{B}) = \text{vec}(\mathbf{A})^T\text{vec}(\mathbf{B})$, the normalization (7) may be rewritten in vector form:

$$\text{vec}(\widetilde{\mathbf{A}}) = \frac{1}{\mathbf{c}^T\text{vec}(\mathbf{A})}\text{vec}(\mathbf{A}), \quad (9)$$

$$\text{where } \mathbf{c} = \frac{1}{m}\text{vec}(\mathbf{\Lambda}^{-1}). \quad (10)$$

We show in section III.1 that $\widetilde{\mathbf{V}}_N$ and $\widetilde{\mathbf{W}}_{N/\sigma_1}$ (with σ_1 of (5)), share the same asymptotic distribution.We start by a lemma needed for proving the main result.

Lemma III.1 *Let $(\mathbf{z}_N)_{N\in\mathbb{N}}$ be a sequence of random vectors such that*

$$\sqrt{N}(\mathbf{z}_N - \mathbf{m}) \xrightarrow{d} \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}). \quad (11)$$

Let us set $\widetilde{\mathbf{z}}_N = \frac{1}{\mathbf{c}^T\mathbf{z}_N}\mathbf{z}_N$ and $\widetilde{\mathbf{m}} = \frac{1}{\mathbf{c}^T\mathbf{m}}\mathbf{m}$, where $\mathbf{c}$ is an arbitrary vector. Then

$$\sqrt{N}(\widetilde{\mathbf{z}}_N - \widetilde{\mathbf{m}}) \xrightarrow{d} \mathcal{N}\left(\mathbf{0}, \mathbf{A}\mathbf{\Sigma}\mathbf{A}^T\right), \quad (12)$$

with $\mathbf{A} = \frac{1}{\mathbf{c}^T\mathbf{m}}\left(\mathbf{I} - \frac{\mathbf{m}\mathbf{c}^T}{\mathbf{c}^T\mathbf{m}}\right)$.

Proof: Let us define $\boldsymbol{\delta}_N = \mathbf{z}_N - \mathbf{m}$. Then

$$\widetilde{\mathbf{z}}_N = \frac{\mathbf{m}+\boldsymbol{\delta}_N}{\mathbf{c}^T(\mathbf{m}+\boldsymbol{\delta}_N)} = \frac{\mathbf{m}+\boldsymbol{\delta}_N}{\mathbf{c}^T\mathbf{m}\left(1+\dfrac{\mathbf{c}^T\boldsymbol{\delta}_N}{\mathbf{c}^T\mathbf{m}}\right)}. \quad (13)$$

Thus, for large N, a first order expansion gives

$$\begin{aligned}\widetilde{\mathbf{z}}_N &= \frac{1}{\mathbf{c}^T\mathbf{m}}\left(1-\frac{\mathbf{c}^T\boldsymbol{\delta}_N}{\mathbf{c}^T\mathbf{m}}\right)(\mathbf{m}+\boldsymbol{\delta}_N)\\ &= \frac{1}{\mathbf{c}^T\mathbf{m}}\left(\mathbf{m}+\boldsymbol{\delta}_N-\frac{\mathbf{c}^T\boldsymbol{\delta}_N}{\mathbf{c}^T\mathbf{m}}\mathbf{m}\right)\\ &= \widetilde{\mathbf{m}}+\mathbf{A}\boldsymbol{\delta}_N.\end{aligned} \tag{14}$$

So, for large N, $\sqrt{N}(\widetilde{\mathbf{z}}_N-\widetilde{\mathbf{m}}) = \mathbf{A}\sqrt{N}\boldsymbol{\delta}_N$, which concludes the proof. ■

Theorem III.1 *Let $\widetilde{\mathbf{V}}_N$ (resp. $\widetilde{\mathbf{W}}_N$) be a $\mathbf{\Lambda}$-normalized M-estimate (resp. Wishart matrix). Then $\sqrt{N}(\widetilde{\mathbf{V}}_N-\mathbf{\Lambda})$ and $\sqrt{\sigma_1 N}(\widetilde{\mathbf{W}}_N-\mathbf{\Lambda})$ share the same asymptotic distribution.*

Proof: From (6) and taking into acount (8), the lemma of section III.1 applied to $\text{vec}(\mathbf{W}_N)$ with $\mathbf{c}$ given by (10) yields

$$\sqrt{\sigma_1 N}(\text{vec}(\widetilde{\mathbf{W}}_N)-\text{vec}(\mathbf{\Lambda})) \xrightarrow{d} \mathcal{N}\left(\mathbf{0},\sigma_1\mathbf{A}\Sigma\mathbf{A}^T\right), \tag{15}$$

$$\begin{aligned}\text{with } \mathbf{A} &= \frac{1}{\mathbf{c}^T\text{vec}(\mathbf{\Lambda})}\left(\mathbf{I}-\frac{\text{vec}(\mathbf{\Lambda})\mathbf{c}^T}{\mathbf{c}^T\text{vec}(\mathbf{\Lambda})}\right)\\ &= \left(\mathbf{I}-\frac{1}{m}\text{vec}(\mathbf{\Lambda})(\text{vec}(\mathbf{\Lambda}^{-1}))^T\right).\end{aligned}$$

Similarly, we obtain from (5),

$$\sqrt{N}(\widetilde{\mathbf{V}}_N-\mathbf{\Lambda}) \xrightarrow{d} \mathcal{N}\left(\mathbf{0},\mathbf{A}'\mathbf{\Pi}\mathbf{A}'^T\right), \tag{16}$$

$$\begin{aligned}\text{with } \mathbf{A}' &= \frac{1}{\mathbf{c}^T\text{vec}(\mathbf{V})}\left(\mathbf{I}-\frac{\text{vec}(\mathbf{V})\mathbf{c}^T}{\mathbf{c}^T\text{vec}(\mathbf{V})}\right)\\ &= \tfrac{1}{\sigma}\mathbf{A},\end{aligned} \tag{17}$$

where the last equality follows from (4). By noticing that

$$\begin{aligned}&\mathbf{A}\text{vec}(\mathbf{V})\\ &= \text{vec}(\mathbf{V})-\frac{1}{m}\text{vec}(\mathbf{V})(\text{vec}(\mathbf{V}^{-1}))^T\text{vec}(\mathbf{V}) = \mathbf{0},\end{aligned} \tag{18}$$

$$\begin{aligned}\text{it follows that } \mathbf{A}'\mathbf{\Pi}\mathbf{A}'^T &= \sigma_1\mathbf{A}'(\mathbf{I}+\mathbf{K})(\mathbf{V}\otimes\mathbf{V})\mathbf{A}'^T\\ &= \sigma_1\mathbf{A}\Sigma\mathbf{A}^T,\end{aligned} \tag{19}$$

which concludes the proof. ■

B. Comments

Performance of signal processing algorithms based on a covariance matrix estimate have been extensively studied in the Gaussian framework with the Wishart distributed SCM. Most of the time, these algorithms are insensitive to a scaling of the covariance matrix estimate, as for instance in DOA estimation. In these cases, the proposed theorem allows a straightforward performance analysis when using M-estimates of $\mathbf{V}$. Indeed for such algorithms, $\mathbf{V}_N$ and $\widetilde{\mathbf{V}}_N$ could be equivalently used, although of course, only $\mathbf{V}_N$ can be obtained in practice. Since theorem 3.2 says that $\widetilde{\mathbf{V}}_N$ and $\widetilde{\mathbf{W}}_N$ share the same distribution for large N, we can conclude that using robust M-estimates is equivalent in terms of performance, to using the SCM with N/σ_1 Gaussian data.

IV. Simulations

To illustrate theorem (III.1), we consider a simulation using the MUltiple SIgnal Classification (MUSIC) method, which estimates the Direction Of Arrival (DOA) of a signal.

A. MUSIC with real data

MUSIC is normally applied to complex baseband signals. However, to highlight the result of theorem (III.1), our analysis of M-estimates has been conducted for real observations. Therefore, we choose to apply the MUSIC algorithm to real narrowband signals. The resulting so called MUSIC-COM algorithm has been described in [16] in a spectral analysis framework with sinusoidal signals; it may be applied to DOA estimation as will be done in this section.

Let $a(t)$ be a white zero-mean Gaussian stationnary signal in the angular frequency band $[\omega_0-\Omega,\omega_0+\Omega]$. $a(t)$ can be written

$$a(t) = s_I(t)\cos(\omega_0 t)-s_Q(t)\sin(\omega_0 t) \tag{20}$$

where $s_I(t)$ and $s_Q(t)$ are independent, white, zero-mean, Gaussian, stationnary signals in $[-\Omega,\Omega]$.

Assuming that $\tau << \frac{\pi}{\Omega}$, we have

$$\begin{aligned}&a(t-\tau)\\ =\;& (s_I(t-\tau)\cos(\omega_0 t)-s_Q(t-\tau)\sin(\omega_0 t))\cos(\omega_0\tau)\\ &+(s_I(t-\tau)\sin(\omega_0 t)+s_Q(t-\tau)\cos(\omega_0 t))\sin(\omega_0\tau)\\ \approx\;& (s_I(t)\cos(\omega_0 t)-s_Q(t)\sin(\omega_0 t))\cos(\omega_0\tau)\\ &+(s_I(t)\sin(\omega_0 t)+s_Q(t)\cos(\omega_0 t))\sin(\omega_0\tau)\end{aligned}$$

where the last approximation follows from $\tau \ll \frac{\pi}{\Omega}$.

$$\text{Let us set: } b(t) = s_I(t)\sin(\omega_0 t)+s_Q(t)\cos(\omega_0 t) \tag{21}$$

$$\text{so that } a(t-\tau)\approx a(t)\cos(\omega_0\tau)+b(t)\sin(\omega_0\tau). \tag{22}$$

It is easy to check that $a(t)$ and $b(t)$ are independent, and that $b(t)$ has the same statistical properties as $a(t)$.

Now let us consider a linear uniform array of M sensors receiving signals from P Gaussian, independent, white, narrowband $[\omega_0-\Omega,\omega_0+\Omega]$, stationary sources. From the above discussion the signals $x_m(t)$ at the sensors output may be written

$$\begin{aligned}&\mathbf{x}(t) = \left(x_1(t)...x_M(t)\right)^T\\ &= \sum_{p=1}^{P} a_p(t)\begin{pmatrix}\cos(\omega_0\tau_{p,1})\\ \vdots\\ \cos(\omega_0\tau_{p,M})\end{pmatrix}+\sum_{p=1}^{P} b_p(t)\begin{pmatrix}\sin(\omega_0\tau_{p,1})\\ \vdots\\ \sin(\omega_0\tau_{p,M})\end{pmatrix}\\ &+\mathbf{n}(t).\end{aligned}$$

where

- $a_p(t)$ and $b_p(t)$ are decorrelated and of same variance σ_p^2,
- $(a_p(t),b_p(t))$ and $(a_{p'}(t'),b_{p'}(t'))$ are decorrelated for $p\neq p'$,
- $\mathbf{n}(t)$ is a white Gaussian noise.
- $\tau_{p,m}$ is the propagation delay between the first and m-th sensors for the p-th signal ($\tau_{p,1}=0$).

Let us sample the array output at a rate $\frac{1}{T_s} = \frac{\Omega}{\pi}$. Let $(\mathbf{x}_1, ..., \mathbf{x}_n, ..., \mathbf{x}_N)$ be N snapshots with $\mathbf{x}_n = \mathbf{x}(nT_s)$:

$$\mathbf{x}_n = \sum_{p=1}^{P} a_{p,n} \begin{pmatrix} \cos(\omega_0 \tau_{p,1}) \\ \vdots \\ \cos(\omega_0 \tau_{p,M}) \end{pmatrix} + \sum_{p=1}^{P} b_{p,n} \begin{pmatrix} \sin(\omega_0 \tau_{p,1}) \\ \vdots \\ \sin(\omega_0 \tau_{p,M}) \end{pmatrix} + \mathbf{n}_n. \quad (23)$$

where

- $a_{p,n} \sim \mathcal{N}(0, \sigma_p^2)$, $b_{p,n} \sim \mathcal{N}(0, \sigma_p^2)$, $\mathbf{n}_n \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$
- $E\left[a_{p,n} a_{p',n'}\right] = E[b_{p,n} b_{p',n'}] = \sigma_p^2 \delta_{p,p'} \delta_{n,n'}$
- $E[a_{p,n} b_{p',n'}] = 0$
- $E[\mathbf{n}_n \mathbf{n}_{n'}^T] = \sigma^2 \delta_{n,n'} \mathbf{I}$

This leads to a real $2P$-dimensional signal subspace for P narrowband sources, instead of the usual P-dimensional subspace. Let $\widehat{\mathbf{M}}$ be the estimated covariance matrix, $\widehat{\mathbf{P}}$ its projector onto the eigenspace related to the $2P$ largest eigenvalues and $\widehat{\mathbf{P}}^{\perp} = \mathbf{I} - \widehat{\mathbf{P}}$. The pseudospectrum of the MUSIC-COM algorithm is

$$F(\theta) = \frac{||\mathbf{a}||^2 + ||\mathbf{b}||^2}{\mathbf{a}^T(\theta)\widehat{\mathbf{P}}^{\perp}\mathbf{a}(\theta) + \mathbf{b}^T(\theta)\widehat{\mathbf{P}}^{\perp}\mathbf{b}(\theta)} \quad (24)$$

with $\mathbf{a}(\theta) = \left(\cos(\omega_0 \tau_1(\theta)) ... \cos(\omega_0 \tau_M(\theta))\right)^T$ and $\mathbf{b}(\theta) = \left(\sin(\omega_0 \tau_1(\theta)) ... \sin(\omega_0 \tau_M(\theta))\right)^T$. $\tau_m(\theta)$ is the theoretical propagation delay between the first and m-th sensor (depending on the distance between two sensors and the angular frequency of the signal) for DOA θ ($\tau_1(\theta) = 0$).
The P largest maxima are reached for θ corresponding to the estimated DOAs of the P arriving signals.

B. Implementation using Huber's M-estimator

We consider $M = 6$ uniform linear array with half wavelength sensors spacing, receiving $P = 1$ Gaussian stationnary narrowband signal with DOA $20°$. The array output is corrupted by an additive spatially white Gaussian noise. SNR per sensor is 5dB, the N snapshots are assumed to be independent.

The employed covariance matrix estimators are the SCM and Huber's M- estimator (example 1, p.53, [6]) defined by

$$u(s) = \frac{1}{\beta} \min\left(1, \frac{k^2}{s}\right) \quad (25)$$

where k^2 and β depend on a parameter q, according to $q = F_m(k^2)$ and $\beta = F_{m+2}(k^2) + k^2 \frac{1-q}{m}$. $F_m(.)$ is the cumulative distribution function of χ^2 variate with m degrees of freedom. Briefly, $q = 1$ leads to the SCM while smaller values bring robustness to outliers. The chosen value is $q = 0.05$ for which $\sigma_1 = 1.3$ in (III.1).

Figure 1 depicts the Mean Square Error (MSE) of the DOA estimated with N data for the SCM and for Huber's estimate. The MSE of the DOA obtained for $\sigma_1 N$ data with Huber's estimate is also represented. For N large enough ($N \geq 20$), this curve and the SCM one overlap, as expected from theorem (III.1).

V. Conclusion

In this paper we have analyzed the statistical properties of general M-estimators of scatter matrix. More precisely, we have shown that these estimators and the classical Wishart matrix share the same asymptotic ditribution, with an appropriate normalization. Indeed, M-estimators have the same asymptotic covariance matrix as the Wishart matrix up to a scalar factor σ_1. Roughly speaking, this means that M-estimators built with $\sigma_1 N$ data achieve the same asymptotic performance as the SCM built with N data. A simulation of DOA estimation using MUSIC method has illustrated this theoretical analysis. These results are in favor of M-estimators since they provide robustness and behave almost like standard tools. The generalisation to the complex case has been done and will be submitted in a forthcoming paper.

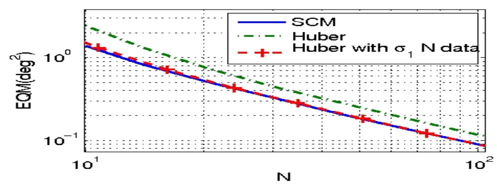

Fig. 1. MSE (logarithmic scale) on the source DOA for Huber's estimate and the SCM.

Acknowledgment

The authors would like to thank the Direction Générale de l'Armement (DGA) to fund this project.

References

[1] S. M. Kay, *Fundamentals of Statistical Signal Processing - Detection Theory*. Prentice-Hall PTR, 1998, vol. 2.
[2] P. J. Huber and E. M. Ronchetti, *Robust statistics*. John Wiley & Sons Inc, 2009.
[3] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel, *Robust Statistics: The Approach Based on Influence Functions*, ser. Wiley Series in Probability and Statistics. John Wiley & Sons, 1986.
[4] R. A. Maronna, D. R. Martin, and J. V. Yohai, *Robust Statistics: Theory and Methods*, ser. Wiley Series in Probability and Statistics. John Wiley & Sons, 2006.
[5] P. J. Huber, "Robust estimation of a location parameter," *The Annals of Mathematical Statistics*, vol. 35, no. 1, pp. 73–101, 1964.
[6] R. A. Maronna, "Robust M-estimators of multivariate location and scatter," *Annals of Statistics*, vol. 4, no. 1, pp. 51–67, January 1976.
[7] M. Bilodeau and D. Brenner, *Theory of multivariate statistics*. Springer Verlag, 1999.
[8] E. Ollila and V. Koivunen, "Robust antenna array processing using M-estimators of pseudo-covariance," in *Proc. 14th IEEE Int. Symp. Personal, Indoor, Mobile Radio Commun.(PIMRC)*, 2003, pp. 7–10.
[9] ——, "Influence functions for array covariance matrix estimators," *Proc. IEEE Workshop on Statistical Signal Processing (SSP),ST Louis, MO*, pp. 445–448, October 2003.
[10] E. Ollila, L. Quattropani, and V. Koivunen, "Robust space-time scatter matrix estimator for broadband antenna arrays," *Vehicular Technology Conference, 2003. VTC 2003-Fall. 2003 IEEE 58th*, vol. 1, pp. 55 – 59 Vol.1, oct. 2003.
[11] F. Gini and M. V. Greco, "Covariance matrix estimation for CFAR detection in correlated heavy tailed clutter," *Signal Processing, special section on SP with Heavy Tailed Distributions*, vol. 82, no. 12, pp. 1847–1859, December 2002.
[12] D. Tyler, "Robustness and efficiency properties of scatter matrices," *Biometrika*, vol. 70, no. 2, p. 411, 1983.
[13] ——, "Some results on the existence, uniqueness, and computation of the M-estimates of multivariate location and scatter," *SIAM Journal on Scientific and Statistical Computing*, vol. 9, p. 354, 1988.
[14] J. T. Kent and D. E. Tyler, "Redescending M-estimates of multivariate location and scatter," *Annals of Statistics*, vol. 19, no. 4, pp. 2102–2119, December 1991.
[15] D. Tyler, "Radial estimates and the test for sphericity," *Biometrika*, vol. 69, no. 2, p. 429, 1982.
[16] P. Stoica and A. Eriksson, "MUSIC estimation of real-valued sine-wave frequencies," *Signal Processing*, vol. 42, no. 2, pp. 139–146, 1995.